Continuity and differentiability of regression M–estimates

María V. Fasano[1] (virfeather@yahoo.com.ar)
Ricardo A. Maronna[2] (rmaronna@retina.ar) Mariela Sued[3] (msued@dm.uba.ar)
Víctor J. Yohai[3] (vyohai@uolsinectis.com.ar)
[1]University of La Plata and CONICET
[2]University of La Plata and C.I.C.P.B.A.
[3]University of Buenos Aires and CONICET

Abstract

This paper deals with the weak continuity, Fisher–consistency and differentiability of estimating functionals corresponding to a class of both linear and nonlinear regression high breakdown M estimates, which includes S and MM estimates. A restricted type of differentiability, called weak differentiability, is defined, which suffices to prove the asymptotic normality of estimates based on the functionals. This approach allows to prove the consistency, asymptotic normality and qualitative robustness of estimates under more general conditions than those required in standard approaches.
Keywords: MM estimates; S estimates; M scale; asymptotic normality; consistency.

1 Introduction

We consider estimation in the regression model with random predictors

$$y_i = g(x_i, \beta_0) + u_i, \tag{1}$$

with data $(x_i, y_i) \in R^p \times R$, $i = 1, ..., n$; where $\beta_0 \in B \subseteq R^q$ is a vector of unknown parameters, $g(x, \beta)$ is a known function continuous in β, and for each i, x_i and u_i are independent. It is assumed that $\{(x_i, y_i), i \geq 1\}$ are identically distributed but not necessarily independent. The well–known fact that the least squares (LS) estimate of β_0 is sensitive to atypical observations has motivated the development of robust estimates.

An important class of robust estimators are the *M estimates.* Inside this class we can distinguish the S estimates introduced by Rousseeuw and Yohai (1984) and the MM estimates proposed by Yohai (1987). For linear regression, S estimates may attain the highest possible breakdown point, and MM estimates may combine the highest possible breakdown point with a high normal efficiency; see e.g. (Maronna, Martin and Yohai (2006), Chapter 5). In the case of nonlinear regression MM estimates may also combine high breakdown point with high normal efficiency. In fact, the normal efficiency of these estimates can be made as close to one as desired, and Monte Carlo simulations in Fasano (2009) show them to have a highly robust behavior for some nonlinear models.

In the nonlinear case, Fraiman (1983) study bounded influence estimates for nonlinear regression. Sakata and White (2001) deal with S estimates for nonlinear regression models with dependent observations; Vainer and Kukush (1998) and Liese and Vajda (2003, 2004) deal with M estimates with fixed scale and therefore no scale equivariant . The latter study the $\sqrt{n}$-consistency of M estimates in more general models, which

include linear and nonlinear regression with independent observations. Stromberg (1995) proved the weak consistency of the least median of squares (LMS) estimate, and Cízek (2005) dealt with the consistency and the asymptotic normality of the least trimmed squares (LTS) estimate under dependency.

Three important qualitative features of these estimates are consistency, asymptotic normality and qualitative robustness. These properties have been studied in the literature through specific approaches. Yohai (1987) proved these properties for MM estimates in the i.i.d. linear case, and Fasano (2009) proved them in the nonlinear case, both assuming symmetrically distributed u_i's.

In this work we propose an alternative approach, based on the representation of the estimates as *functionals* on distributions (Hampel 1971). For a large class of estimates, which includes M estimates, one can define a functional $T(G)$ on the space of data distributions, such that if G_n is the empirical distribution, then $T(G_n)$ is the estimate, and if G_0 is the underlying distribution, then $T(G_0)$ is the parameter that we want to estimate. The weak continuity of the functional T simplifies the proof of consistency of $T(G_n)$ and some suitable forms of differentiability of T, as Fréchet or Hadamard differentiability, allow simple proofs of the asymptotic normality of the estimate under very general conditions. These results hold without the requirement that G_n be the empirical distribution of a sequence of i.i.d. random variables: if we want to estimate $T(G_0)$, it suffices that G_n converges weakly to G_0 a.s.. The weak continuity of M functionals at a general statistical model were studied by Clarke (1983 and 2000). Fréchet differentiability was studied by Boos and Serfling (1980) and Clarke (1983), and Hadamard differentiability by Fernholz (1983). In all of these works, it is required that the score function used for the M estimate be bounded, and therefore their results can not be applied to regression M estimates. In this paper we prove under very general conditions that the functionals associated to M estimates of regression are weakly continuous. Besides, since the usual forms of differentiability , like Fréchet or Hadamard differentiability, require in the case of M estimates the boundedness of the score functions, we introduce a new concept of differentiability, that we call weak differentiability, which is satisfied for high breakdown M estimates of regression, e.g., by S and MM estimates, and which is adequate to prove the asymptotic normality of these estimates.

This work is organized as follows: In Section 2 we define the estimates to be considered and in Sections 3, 4 and 5 we shall respectively deal with the continuity, the Fisher–consistency, and differentiability of the functionals corresponding to the estimates defined above. These results will be shown to imply the consistency, qualitative robustness and asymptotic normality of the estimates under assumptions more general than the i.i.d. model and without the requirement of symmetric errors. In Section 6 we apply the results obtained in the former Sections to MM estimates. Finally Section 7 contains all proofs.

2 Definitions of estimates

We first define our notation. Henceforth $\mathrm{E}_G[h(z)]$ and $\mathrm{P}_G(A)$ will respectively denote the expectation of $h(z)$ and the probability that $z \in A$, when z is distributed according to G. If z has distribution G we write $z \sim G$ or $\mathcal{D}(z) = G$. Weak convergence of distributions, convergence in probability and convergence in distribution of random variables or vectors are denoted by $G_n \to_w G$, $z_n \to_p z$ and $z_n \to_d z$, respectively. By an abuse of notation, we will write $z_n \to_d G$ to denote $\mathcal{D}(z_n) \to_w G$. The complement and the indicator of the set A are denoted by A^c and $\mathbf{1}_A$, respectively. The scalar product of vectors a and b is denoted by $a'b$, and R_+ denotes the set of positive real numbers.

Before proceeding further, we need to clarify an important detail. If it is not assumed that the errors have a symmetric distribution, the standard treatment of regression estimates requires some condition related to the "centering" of the u_i to ensure the identifiability of all parameters and the consistency of the estimates. For LS, this condition is $\mathrm{E}u_i = 0$. For M estimates it is $\mathrm{E}\psi(u_i/\sigma) = 0$, where ψ is the score function and σ is an error scale; the fact that this assumption depends on σ, which is an unknown parameter, makes

it undesirable. Since we want our results to hold under more general assumptions, we will employ another (somewhat nonstandard) approach to identify β_0. Note first that in the linear case, if there is a constant term, the slopes are always identifiable no matter the distribution of u_i, but the intercept is unidentified without some centering assumption on u_i, such as zero median. For these reasons, besides β_0, our M estimates will include an additional additive term α. If the model does contain an intercept, then α will single it out, and g will be redefined as the "non–intercept"part of the model. Otherwise, α may be interpreted as a "centering constant" for u_i. In general, α remains unidentified; if it has to be identified (e.g. for prediction) then some assumption on the centering of u_i must be added to the model.

Instead of a centering condition we will require the following identifiability condition:

$$\mathrm{P}(g(x_i,\beta_0) = g(x_i,\beta) + \alpha) < 1 \ \ \forall\ \beta \neq \beta_0,\ \forall\ \alpha. \tag{2}$$

Otherwise model (1) might also hold with β instead of β_0 and $u_i + \alpha$ instead of u_i. In the linear case $g(x,\beta) = \beta' x$ this condition means that g does not include an intercept and x_i is not concentrated on any hyperplane.

Now in order to get consistent estimators for β_0 our estimates must always contain a term which plays the role of an intercept. Let henceforth $\xi = (\beta', \alpha)'$ with $\alpha \in R$, and define the function

$$\underline{g}(x,\xi) = g(x,\beta) + \alpha.$$

M estimates are then defined as

$$\widehat{\xi}_{\mathrm{M}} = \arg\min_{\xi\in B\times R} \sum_{i=1}^{n} \rho\left(\frac{y_i - \underline{g}\,(x_i,\xi)}{\widehat{\sigma}}\right), \tag{3}$$

where $\widehat{\sigma}$ is a robust residual scale and ρ is a loss function.

To define *S estimates* we need an M scale $S(r)$. Given $r = (r_1, ..., r_n)'$, $S(r)$ is defined as the solution σ of

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\frac{r_i}{\sigma}\right) = \delta, \tag{4}$$

where ρ_0 is another loss function and the constant δ regulates the estimate's robustness.

Then, S estimates of regression are defined by

$$\widehat{\xi}_{\mathrm{S}} = \arg\min_{\xi\in B\times R} S\left(r\left(\xi\right)\right), \tag{5}$$

where $r(\xi)$ is the residual vector with elements $r_i(\xi) = y_i - \underline{g}(x_i,\xi)$.

In particular we will consider with some detail the subclass of MM estimates. These estimates are defined by (3) with $\widehat{\sigma}$ obtained from an S estimate, namely

$$\widehat{\sigma} = \min_{\xi\in B\times R} S\left(r\left(\xi\right)\right) \tag{6}$$

with $\rho \leq \rho_0$. Yohai (1987) showed that in case of linear regression the asymptotic breakdown point of MM estimates with $\delta = 0.5$ is 0.5 if $\mathrm{P}(\beta' x_i + a = 0) = 0$ for all $\beta \neq 0$, and that, simultaneously, it is possible to choose ρ so that the corresponding MM estimate yields an arbitrarily high efficiency when the errors are Gaussian.

Now in order to state our results, we must first express the already defined M and S estimates as functionals. Throughout this article loss functions will be bounded ρ-functions, in the following sense.

Definition 1 *A* bounded ρ–function *is a function* $\rho(t)$ *that is a continuous nondecreasing function of* $|t|$, *such that* $\rho(0)=0$, $\rho(\infty)=1$, *and* $\rho(v)<1$ *implies that* $\rho(u)<\rho(v)$ *for* $|u|<|v|$.

Then, in the rest of the paper we will assume the following property

R0. ρ and ρ_0 are "bounded ρ–functions."

Define the residual scale functional $S^*(G,\xi)$ by

$$\mathrm{E}_G\rho_0\left(\frac{y-\underline{g}(x,\xi)}{S^*(G,\xi)}\right)=\delta, \tag{7}$$

for $\delta\in(0,1)$. Then the regression S functional T_{S} and the associated error scale M functional $S(G)$ are respectively defined by

$$T_{\mathrm{S}}(G):=(T_{\mathrm{S},\beta}(G),T_{\mathrm{S},\alpha}(G))=\arg\min_{\xi\in B\times R}S^*(G,\xi) \tag{8}$$

and

$$S(G)=\min_{\xi\in B\times R}S^*(G,\xi). \tag{9}$$

We will deal with a regression M functional $T_{\mathrm{M}}(G)$ defined as

$$T_{\mathrm{M}}(G):=(T_{\mathrm{M},\beta}(G),T_{\mathrm{M},\alpha}(G))=\arg\min_{\xi\in B\times R}M_G(\xi), \tag{10}$$

where the function $M_G:B\times R\to R$ is

$$M_G(\xi)=\mathrm{E}_G\rho\left(\frac{y-\underline{g}(x,\xi)}{\widetilde{S}(G)}\right) \tag{11}$$

and $\widetilde{S}(G)$ is an arbitrary residual scale functional, for example the one defined in (9).

It is easy to show that the S regression functional defined in (8) is also an M functional. In fact $T_{\mathrm{S}}(G)$ coincides with $T_{\mathrm{M}}(G)$ when in (11) we have $\rho=\rho_0$ and $\widetilde{S}(G)=S(G)$. We may then write

$$T_{\mathrm{S}}(G)=\arg\min_{\xi\in B\times R}\mathrm{E}_G\rho_0\left(\frac{y-\underline{g}(x,\xi)}{S(G)}\right). \tag{12}$$

Remark 1 *In general, the minimum at (8) or (10) might be attained at more than one value of* ξ. *It will be henceforth assumed that the functional is well-defined by the choice of a single value. Our results will not depend on how the choice is made. However, it will be shown in Section 4 that under very general conditions, if* G_0 *is the distribution of* (x,y) *satisfying (1), then* $T_{\mathrm{S}}(G_0)$ *and* $T_{\mathrm{M}}(G_0)$ *are unique and* $T_{\mathrm{S},\beta}(G_0)=T_{\mathrm{M},\beta}(G_0)=\beta_0$ *(Fisher–consistency).*

3 Weak continuity of M and S regression functionals

We will show the weak continuity of the functionals defined above in two cases: nonlinear regression with a compact parameter space B, and linear regression.

Define for $G=\mathcal{D}(x,y)$

$$c(G)=\sup\{\mathrm{P}_G(\beta'x+\alpha=0):\beta\in R^p,\alpha\in R,\beta\neq 0\}. \tag{13}$$

Theorem 1 *Let $G_0 = \mathcal{D}(x, y)$ be such that (10) has a unique solution $T_{\mathrm{M}}(G_0)$. Assume that $\widetilde{S}$ is weakly continuous at G_0 and $\widetilde{S}(G_0) > 0$. Then $T_{\mathrm{M}} = (T_{\mathrm{M},\beta}, T_{\mathrm{M},\alpha})$ is weakly continuous at G_0 if either (a) B is compact, or (b) $B = R^p$, $g(x, \beta) = \beta' x$ and*

$$M_{G_0}(T_{\mathrm{M}}(G_0)) < 1 - c(G_0). \tag{14}$$

Theorem 2 *Let $G_0 = \mathcal{D}(x, y)$ be such that $T_{\mathrm{S}}(G_0)$ is unique and $S(G_0) > 0$. Assume that either (a) B is compact, or (b) $B = R^p$, g is linear, i.e., $g(x, \beta) = \beta' x$ and $\delta < 1 - c(G_0)$ with $c(G)$ defined in (13). Then $S(G)$ and $T_{\mathrm{S}}(G) = (T_{\mathrm{S},\beta}, T_{\mathrm{S},\alpha})$ are weakly continuous at G_0.*

Let now G_0 be the distribution of (x, y) under model (1), and assume that T_{M} (respectively T_{S}) is Fisher–consistent for β_0, i.e., $T_{\mathrm{M},\beta}(G_0) = \beta_0$ (respectively $T_{\mathrm{S},\beta}(G_0) = \beta_0$). Then the former results imply that $T_{\mathrm{M},\beta}$ (respectively $T_{\mathrm{S},\beta}$) evaluated at the empirical distribution is consistent whenever the empirical distributions converge to the underlying one. More precisely, we have the following result:

Corollary 1 *Assume the same hypotheses as in Theorem 1 (respectively Theorem 2) plus the Fisher–consistency of T_{M} (respectively T_{S}): $T_{\mathrm{M},\beta}(G_0) = T_{\mathrm{S},\beta}(G_0) = \beta_0$. Call G_n the empirical distribution of $\{(x_i, y_i) : i = 1, ..., n\}$. If $G_n \to_w G_0$ a.s., then $\{T_{\mathrm{M},\beta}(G_n)\}$ (respectively $\{T_{\mathrm{S},\beta}(G_n)\}$) is strongly consistent for β_0.*

This result is immediate. The a.s. weak convergence of G_n to G_0 is well–known to hold for i.i.d. (x_i, y_i) (see e.g. (Billingsley 1999, Problem 3.1)). It holds also under more general assumptions on the joint distribution of $\{(x_i, y_i) : i \geq 1\}$, such as ergodicity.

We now turn to qualitative robustness. Consider a sequence of estimates $\{\widehat{\xi}_n\}$ based on a functional T, i.e. $\widehat{\xi}_n = T(G_n)$ where G_n is the empirical distribution corresponding to data $(z_1, ..., z_n)$. Hampel (1971) proved that for $\{\widehat{\xi}_n\}$ to be qualitatively robust at a distribution G_0 it suffices that T be weakly continuous at G_0 and $\widehat{\xi}_n$ be a continuous function of $(z_1, ..., z_n)$,

Papantoni-Kazakos and Grey (1979) employ a weaker definition of robustness, which they call *asymptotic qualitative robustness*, and prove that it is equivalent to weak continuity. Therefore Theorems 1 and 2 imply the asymptotic qualitative robustness of T_{M} and T_{S}.

4 Fisher–consistency of M and S estimates

In this Section we give sufficient conditions to guarantee that both (8) and (10) are minimized at unique values, and to guarantee the Fisher consistency for β_0.

Recall that a density f is *strongly unimodal* if there exists a such that $f(t)$ is nondecreasing for $t < a$, nonincreasing for $t > a$, and f has a unique maximum at $t = a$.

Theorem 3 is an auxiliary result, which is a small variation of one given by Mizera (1993). We will need the following condition on ρ

R1. For some m, $\rho(u) = 1$ iff $|u| \geq m$, and $\log(1 - \rho)$ is concave on $(-m, m)$

Theorem 3 *Let ρ satisfy Condition R1 and let F be a distribution with a strongly unimodal density f. Then (a) there exists t_0 such that*

$$q(t) = \mathrm{E}_F \rho(u - t) \tag{15}$$

has a unique minimum at t_0; (b) if F is symmetric around μ_0, then $t_0 = \mu_0$.

It is easy to check that condition R1 with $m = k$ holds in particular for the popular family of bisquare functions, defined by

$$\rho_k\left(u\right) = 1 - \left(1 - \left(\frac{u}{k}\right)^2\right)^3 I(|u| \leq k).$$

We will establish the Fisher–consistency of T_{M}. Put for brevity $\sigma = S\left(G_0\right)$ and let F_0 be the distribution of u_i in (1) and assume that it has a strongly unimodal density. Let Δ denotes the unique minimizer of $\mathrm{E}_{F_0}\rho((u-t)\,/\sigma)$; note that if u_i is symmetric around μ_0, then part (b) of Theorem 3 implies that $\Delta = \mu_0$.

Theorem 4 *Let G_0 be the joint distribution of (x_i, y_i) satisfying model (1), where u_i has distribution F_0 with a strongly unimodal density. Assume that the identifiability condition (2) and condition R1 hold. Then, $M_{G_0}(\xi)$ is minimized at the unique point $T_{\mathrm{M}}\left(G_0\right) = (\beta_0, \Delta)$, and so T_{M} is Fisher–consistent for β_0, i.e., $T_{M,\beta}(G_0) = \beta_0$. If we also assume that F_0 is symmetric around μ_0, we have $T_{M,\alpha}(G_0) = \mu_0$.*

Remark 2 *Theorem 4 gives also sufficient conditions for the Fisher–consistency of the regression S functional T_{S} . In fact, according to (12), T_{S} is also an M functional.*

5 Differentiability of estimating functionals

In this Section we shall first deal with the differentiability of general functionals and then specialize to our regression case. Let $\mathcal{G}_h$ be a set of distributions on R^h. Consider an estimating functional $T : \mathcal{G}_h \to R^k$. Hampel (1976) defines the *influence function* of T at $G \in \mathcal{G}_h$ as the function $I_{T,G}(z) : R^h \to R^k$

$$I_{T,G}(z) = \left.\frac{\partial(T((1-\varepsilon)G + \varepsilon\delta_z)}{\partial\varepsilon}\right|_{\varepsilon=0}, \tag{16}$$

where δ_z is the point mass distribution at z. Given a distance d on $\mathcal{G}_h$ which metricizes the topology of convergence in distribution, T is *Fréchet differentiable* at G_0 under d if

$$T(G) - T(G_0) = \mathrm{E}_G I_{T,G_0}(z) + o(d(G, G_0)).$$

Fréchet differentiability can be used to prove the asymptotic normality of the estimate. However, Fréchet differentiability also requires that $I_{T,G}(z)$ be bounded. Since this condition is not satisfied by regression M estimates, we are going to define a weaker type of differentiability, which suffices to prove asymptotic normality.

Definition 2 *Let T be an estimating functional that is weakly continuous at G_0, and consider a sequence $\{G_n\}$ such that $G_n \to_w G_0$. We say that T is* weakly differentiable at $\{G_n\}$ *if*

$$T(G_n) - T(G_0) = \mathrm{E}_{G_n} I_{T,G_0}(z) + o\left(\|\mathrm{E}_{G_n} I_{T,G_0}(z)\|\right). \tag{17}$$

The definition of weak differentiability helps understanding the asymptotic behavior of $T(G_n) - T(G_0)$, as the next Lemma shows.

Lemma 1 *Consider a random sequence of distributions $\{G_n\}$ converging weakly to G_0 a.s. Suppose that T is weakly differentiable at $\{G_n\}$ a.s. and that for some sequence $\{a_n\}$ of real numbers*

$$a_n \mathrm{E}_{G_n} I_{T,G_0}(z) \to_d H.$$

Then

$$a_n(T(G_n) - T(G_0)) = a_n \mathrm{E}_{G_n} I_{T,G_0}(z) + o_p(1). \tag{18}$$

and therefore $a_n(T(G_n) - T(G_0)) \to_d H$ too.

The proof of this Lemma is immediate.

Remark. Note that if (18) holds for a joint functional $T=(T_1,T_2)$, it also holds for T_1, i.e.,

$$a_n(T_1(G_n)-T_1(G_0))=a_n\mathrm{E}_{G_n}I_{T_1,G_0}(z)+o_p(1). \tag{19}$$

We now deal with the differentiability of a *general M estimating functional*, i.e., a functional T defined on a subset of $\mathcal{G}_p$ with values in R^q, that for some function $\Psi: R^p\times R^q\to R^q$ satisfies the equation

$$\mathrm{E}_G\Psi(z,T(G))=0. \tag{20}$$

We will assume that Ψ is continuously differentiable with respect to θ and call $\dot{\Psi}(z,\theta)$ (or alternatively $\partial\Psi(z,\theta)/\partial\theta$) the $q\times q$ differential matrix with elements $\dot{\Psi}_{jk}(z,\theta)=\partial\Psi_j(z,\theta)/\partial\theta_k$. Define

$$D(G,\theta)=\mathrm{E}_G\dot{\Psi}(z,\theta). \tag{21}$$

Let $\theta_0=T(G_0)$ and assume that

$$D_0=D(G_0,\theta_0) \tag{22}$$

exists. Suppose that D_0 is nonsingular, that T is weakly continuous at G_0 and that there exists $\eta>0$ such that

$$\mathrm{E}_{G_0}\sup_{||\theta-\theta_0||\leq\eta}||\dot{\Psi}(z,\theta)||<\infty, \tag{23}$$

where $\|.\|$ denotes the l_2 norm. Then, it is easy to show that the influence function of T at G_0 is given by

$$I_{T,G_0}(z)=-D_0^{-1}\Psi(z,\theta_0). \tag{24}$$

The following conditions are sufficient for the weak differentiability of T at $\{G_n\}$.

Condition 1 *$\{G_n\}$ is a sequence of distribution functions that converges weakly to G_0 and*

$$\lim_{\eta\to0}\limsup_{n\to\infty}\sup_{\|\theta-\theta_0\|\leq\eta}\|D(G_n,\theta)-D_0\|=0. \tag{25}$$

Condition 2 *$\{G_n\}$ is a sequence of distribution functions such that, at a neighborhood of θ_0, for each n*

$$\frac{\partial}{\partial\theta}\mathrm{E}_{G_n}\Psi(z,\theta)=\mathrm{E}_{G_n}\frac{\partial}{\partial\theta}\Psi(z,\theta). \tag{26}$$

Theorem 5 *Assume that T is an M functional satisfying (20) and weakly continuous at G_0, that $\dot{\Psi}(z,\theta)$ is continuous in θ, D_0 is non singular and there exists $\eta>0$ such that (23) holds. Suppose that $\{G_n\}$ satisfies Condition 1 and Condition 2; then T is weakly differentiable at $\{G_n\}$.*

The following Theorem gives sufficient conditions for a.s. differentiability of M functionals, at a random sequence of distributions. In particular, it includes the case where G_n are the empirical distributions of observations corresponding to an ergodic process.

Theorem 6 *Let $\{G_n\}$ be a sequence of random distribution converging weakly to G_0 and satisfying Condition 2 a.s.. Assume also that $\dot{\Psi}(z,\theta)$ is continuous in θ, that there exists $\eta>0$ such that (23) holds and that D_0 is nonsingular. Let T be an M functional satisfying (20) and weakly continuous at G_0. Then T is weakly differentiable at $\{G_n\}$ a.s. in any of the following two cases: (a) for each function $d(z)$ such that $E_{G_0}|d(z)|<\infty$, on a set of probability one we have that $\{\mathrm{E}_{G_n}d(z)\}$ converges to $\mathrm{E}_{G_0}d(z)$, or (b) $\dot{\Psi}(z,\theta)$ is bounded.*

Corollary 2 *Let $\{G_n\}$ be a sequence of empirical distributions associated to i.i.d. $\{z_i\}$ with distribution G_0. Assume that $\dot{\Psi}(z,\theta)$ is continuous in θ, that there exists $\eta>0$ such that (23) holds, that D_0 is nonsingular and that $I_{T,G_0}(z)$ has finite second moments under G_0. Let T be an M functional continuous at G_0. Then $n^{1/2}(T(G_n)-T(G_0))\rightarrow_d N(0,V)$ with*

$$V=\mathrm{E}_{G_0}I_{T,G_0}(z)\,I_{T,G_0}(z)'. \tag{27}$$

6 MM estimates

In this Section we will summarize the properties derived from Theorems 1-6 for S and MM estimates of regression and location.

6.1 Regression case

Recall that MM estimates, which we denote here by $T_{\mathrm{MM}}=(T_{\mathrm{MM},\beta},T_{\mathrm{MM},\alpha})$, are defined in (10), where $\widetilde{S}$ is the functional S defined in (9) with $\rho_1\leq\rho_0$, where we use ρ_1 to denote the ρ-function employed in (11). As mentioned above, the definition of $\widehat{\xi}_{\mathrm{MM}}$ in (3) requires also $\widehat{\sigma}$ defined by (6), and hence also $\widehat{\xi}_{\mathrm{S}}$ defined in (5). Therefore, these three estimates must be considered simultaneously. Call

$$\widehat{\theta}=\left(\widehat{\xi}_{\mathrm{S}},\widehat{\xi}_{\mathrm{MM}},\widehat{\sigma}\right) \tag{28}$$

the joint solution of (3)-(5)-(6).

In the remaining of this Section we assume the following properties:

R2. ρ_0 and ρ_1 are twice continuously differentiable

We denote by ψ_0 and ψ_1 the derivatives of ρ_0 and ρ_1, respectively. Assume also that $g(x,\beta)$ satisfies

R3 g is twice continuously differentiable with respect to β.

We denote by $\underline{\dot{g}}(x,\xi)$ and $\underline{\ddot{g}}(x,\xi)$ the vector of first derivatives and the matrix of second derivatives of $\underline{g}$ with respect to ξ, respectively. Analogously we denote by $\dot{g}(x,\beta)$ and $\ddot{g}(x,\beta)$ the vector of first derivatives and the matrix of second derivatives of g with respect to β, respectively.

Differentiating (3) we have that $\widehat{\xi}_{\mathrm{MM}}$ satisfies the system

$$\frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\frac{y_i-\underline{g}(x_i,\widehat{\xi}_{\mathrm{MM}})}{\widehat{\sigma}}\right)\underline{\dot{g}}(x_i,\widehat{\xi}_{\mathrm{MM}})=0. \tag{29}$$

It is immediate that $\widehat{\xi}_{\mathrm{S}}$ also satisfies

$$\widehat{\xi}_{\mathrm{S}}=\arg\min_{\xi\in B\times R}\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{y_i-\underline{g}(x_i,\xi)}{\widehat{\sigma}}\right).$$

Then, differentiating this equation we get

$$\frac{1}{n}\sum_{i=1}^{n}\psi_0\left(\frac{y_i-\underline{g}(x_i,\widehat{\xi}_{\mathrm{S}})}{\widehat{\sigma}}\right)\underline{\dot{g}}(x_i,\widehat{\xi}_{\mathrm{S}})=0. \tag{30}$$

Finally according to (4), $\widehat{\sigma}$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{y_i-\underline{g}(x_i,\widehat{\xi}_{\mathrm{S}})}{\widehat{\sigma}}\right)-\delta=0. \tag{31}$$

Then $\widehat{\theta}$ satisfies the system of $2q+3$ equations (29)-(30)-(31). Putting $z_i = (x_i, y_i)$ and denoting by G_n the empirical distribution of $\{z_1, .., z_n\}$, this system can be written as

$$\frac{1}{n}\sum_{i=1}^{n}\Psi\left(z_i,\widehat{\theta}\right) = \mathrm{E}_{G_n}\Psi\left(z,\widehat{\theta}\right) = 0, \tag{32}$$

where if $\theta = (\xi_{\mathrm{S}}, \xi_{\mathrm{MM}}, \sigma)$, $\Psi(z,\theta)$ is defined by

$$\Psi\left(z,\theta\right) = \left[\begin{array}{c} \psi_0\left(\frac{y-\underline{g}(x,\xi_{\mathrm{S}})}{\sigma}\right)\underline{\dot{g}}(x,\xi_{\mathrm{S}}) \\ \psi_1\left(\frac{y-\underline{g}(x,\xi_{\mathrm{MM}})}{\sigma}\right)\underline{\dot{g}}(x,\xi_{\mathrm{MM}}) \\ \rho_0\left(\frac{y-\underline{g}(x,\xi_{\mathrm{S}})}{\sigma}\right)-\delta. \end{array}\right].$$

Let

$$T(G) = \left(T_{\mathrm{S}}\left(G\right), T_{\mathrm{MM}}\left(G\right), S\left(G\right)\right) \tag{33}$$

be the estimating functional associated to $\widehat{\theta}$. Then, if (23) holds, we can differentiate the functions to be minimized in (10) and (12) inside the expectation, obtaining that $T(G)$ satisfies the equation

$$\mathrm{E}_G\Psi(z, T(G)) = 0. \tag{34}$$

Note that the solution to this equation is in general not unique, and therefore, T is not defined exclusively by this equation.

To verify (23), in addition to R0-R3 we need the following assumption:

R4. For some $\eta > 0$

$$\mathrm{E}\sup_{\|\beta-\beta_0\|\leq\eta}\|\dot{g}(x,\beta)\|^2 < \infty \ \text{ and } \ \mathrm{E}\sup_{\|\beta-\beta_0\|\leq\eta}\|\ddot{g}(x,\beta)\| < \infty. \tag{35}$$

Suppose that D_0 defined by (22) is non singular, then under these assumptions, we also get that $I_{T,G_0}(z)$ has finite second moments under G_0. Note that in the case of linear regression, (35) reduces to $\mathrm{E}_{G_0}\|x\|^2 < \infty$.

Define

$$\alpha_{0i} = \arg\min_t \mathrm{E}_{F_0}\rho_i\left(\frac{u-t}{S(G_0)}\right), \ i = 0, 1, \tag{36}$$

where F_0 is the distribution of u_i in model (1). We will see in Theorem 7 that under some general conditions, $T_{\mathrm{S},\alpha}(G_0) = \alpha_{00}$ and $T_{\mathrm{MM},\alpha}(G_0) = \alpha_{01}$..

Put $\theta_0 = (\beta_0, \alpha_{00}, \beta_0, \alpha_{01}, \sigma_0)$ with $\sigma_0 = S(G_0)$.The following numbers, vectors and matrices are required to derive a closed formula for the influence functions of T_{MM} and T_{S}. Let

$$a_{0i} = \mathrm{E}_{G_0}\psi_i'\left(\frac{y-g(x,\beta_0)-\alpha_{0i}}{\sigma_0}\right) = \mathrm{E}_{F_0}\psi_i'\left(\frac{u-\alpha_{0i}}{\sigma_0}\right), \ i = 0, 1,$$

$$e_{0i} = \mathrm{E}_{F_0}\left(\frac{u-\alpha_{0i}}{\sigma_0}\right)\psi_0'\left(\frac{u-\alpha_{0i}}{\sigma_0}\right), \ i = 0, 1,$$

$$d_0 = \mathrm{E}_{F_0}\left(\frac{u-\alpha_{00}}{\sigma_0}\right)\psi_0\left(\frac{u-\alpha_{00}}{\sigma_0}\right),$$

$$b_0 = \mathrm{E}_{G_0}\dot{g}(x,\beta_0), \quad b_0^* = (b_0', 1)',$$

$$A_0 = \mathrm{E}_{F_0}(\dot{g}(x,\beta_0) - b_0)(\dot{g}(x,\beta_0) - b_0)'$$

and

$$C_0 = \begin{bmatrix} A_0 + b_0 b_0' & b_0 \\ b_0' & 1 \end{bmatrix}. \tag{37}$$

It is shown in Section 7.5 that the influence function of T_{MM} is given by

$$I_{T_{\mathrm{MM},\beta},G_0}(x,y) = \frac{\sigma_0}{a_{01}} \psi_1\left(\frac{y - \underline{g}(x,(\beta_0,\alpha_{01}))}{\sigma_0}\right) A_0^{-1}\left(\dot{g}(x,\beta_0) - b_0\right) \tag{38}$$

and

$$\begin{aligned} I_{T_{\mathrm{MM},\alpha},G_0}(x,y) = & -\frac{\sigma_0}{a_{01}}\psi_1\left(\frac{y - \underline{g}(x,(\beta_0,\alpha_{01}))}{\sigma_0}\right)\left[1 + b_0' A_0^{-1}\left(b_0 - \dot{g}(x,\beta_0)\right)\right] \\ & + \frac{\sigma_0 e_{01}}{a_{01} d_0}\left(\rho_0\left(\frac{y - \underline{g}(x,(\beta_0,\alpha_{01}))}{\sigma_0}\right) - \delta\right). \end{aligned} \tag{39}$$

The influence functions of $T_{\mathrm{S},\beta}$ and $T_{\mathrm{S},\alpha}$ can be obtained similarly replacing α_{01},a_{01} and e_{01} by α_{00},a_{00} and e_{00} respectively.

If the errors u_i have a symmetric distribution F_0, then $e_{01} = 0$ and $\alpha_{01} = \alpha_{00} = \alpha_0$, the center of symmetry of F_0..This entails a considerable simplification of the influence function $I_{T_{\mathrm{MM}}}$. In fact, in this case we get

$$I_{T_{\mathrm{MM},},G_0}(z) = \frac{\sigma_0}{\mathrm{E}_{F_0}\psi_1'\left((u-\alpha_0)/\sigma_0\right)}\psi_1\left(\frac{y - g(x,\beta_0) - \alpha_0}{\sigma_0}\right) C_0^{-1}\underline{\dot{g}}(x,\beta_0) ,$$

and the asymptotic covariance matrix (27) is

$$V = \sigma_0^2 \frac{\mathrm{E}_{F_0}\psi_1\left((u-\alpha_0)/\sigma_0\right)^2}{\left(\mathrm{E}_{F_0}\psi_1'\left((u-\alpha_0)/\sigma_0\right)\right)^2} C_0^{-1}. \tag{40}$$

The next Theorem 7 summarizes the properties of S and MM regression functionals

Theorem 7 *Let $z = (x,y)$ satisfy model (1) where the distribution F_0 of u_i has a strong unimodal density and the identifiability condition (2) holds. Assume that ρ_0 and ρ_1 are bounded ρ-functions that satisfy R1, with $\rho_1(u) \leq \rho_0(u)$. Let T be defined by (33) and G_0 the distribution of (x,y). Then, we have:*

(i) $T_{\mathrm{S}}(G_0) = (\beta_0, \alpha_{00})$ is the unique minimizer in (8). If F_0 is symmetric with respect to μ_0 we have $\alpha_{00} = \mu_0$.

(ii) $T_{\mathrm{MM}}(G_0) = (\beta_0, \alpha_{01})$ is the unique minimizer in (10). If F_0 is symmetric with respect to μ_0 we have $\alpha_{01} = \mu_0$.

(iii) The functional $T = (T_{\mathrm{S}}, T_{\mathrm{MM}}, S)$ is weakly continuous at G_0 if either (a) B is compact, or (b) $B = R^p$, $g(x,\beta) = \beta' x$ and $\delta < 1 - c(G_0)$.

(iv) Assume also that R2, R3, R4 hold, that $a_{00} \neq 0$, $a_{01} \neq 0$, $d_0 \neq 0$ and that A_0 is invertible. Then, $D_0 = \mathrm{E}_{G_0}\dot{\Psi}\left(z, T(G_0)\right)$ is invertible, $I_{T_{\mathrm{MM},\beta},G_0}(x,y)$ and $I_{T_{\mathrm{MM},\alpha},G_0}(x,y)$ are given by (38) and (39), respectively, while the influence functions $I_{T_{\mathrm{MM},\beta},G_0}(x,y)$ and $I_{T_{\mathrm{MM},\alpha},G_0}(x,y)$ have a similar expression replacing α_{01},a_{01} and e_{01} by α_{00},a_{00} and e_{00}, respectively.

(v) Under the same assumptions as in (iv), let $\{G_n\}$ be a sequence of random distributions converging weakly to G_0 and satisfying Condition 2 a.s.. Suppose also that for each function $d(z)$ such that $\mathrm{E}_{G_0}|d(z)| < \infty$, we have that $\{\mathrm{E}_{G_n} d(z)\}$ converges to $\mathrm{E}_{G_0} d(z)$ a.s.. Then, the functional T is weakly differentiable at $\{G_n\}$.

(vi) Assume the same conditions as in (iv) and that

$$n^{1/2} I_{T,G_0}(G_n) \to_d H, \tag{41}$$

Then

$$n^{1/2}(T(G_n) - T(G_0)) = n^{1/2} I_{T,G_0}(G_n) + o_p(1) \tag{42}$$

and therefore

$$n^{1/2}(T(G_n) - T(G_0)) \to_d H. \tag{43}$$

(vii) Assume the same conditions as in (iv). Let G_n be the sequence of empirical distributions corresponding to i.i.d. observations $\{(x_i, y_i) : i \geq 1\}$ with common distribution G_0. Then (41) holds with $H = N(0, V)$ and V given by (27).

6.2 Location case

The location model corresponds to the case where there are no regressors: $p = q = 0$ and so $y_i = u_i$ and $\xi = \alpha$. If F_0 denotes the common distribution of the u_i, then $T(F_0) = (T_{\mathrm{S}}(F_0), T_{\mathrm{MM}}(F_0), S(F_0))$ is defined as in the regression case with $\underline{g}(x, \xi)$ replaced by α. Then, the resulting $T_{\mathrm{MM}} = T_{\mathrm{MM},\alpha}$ and $T_{\mathrm{S}} = T_{\mathrm{S},\alpha}$ are the location functionals while $\overline{S}$ is a functional estimating the error scale. In this case, I_{T_{MM},F_0} is given by

$$\begin{aligned} I_{T_{\mathrm{MM}},F_0}(x) = {} & \frac{\sigma_0}{a_{01}} \psi_1 \left(\frac{y - \alpha_{01}}{\sigma_0} \right) \\ & - \frac{e_{01}\sigma_0}{a_{01} d_0} \left(\rho_0 \left(\frac{y - \alpha_{00}}{\sigma_0} \right) - \delta \right). \end{aligned} \tag{44}$$

The following Theorem summarizes the properties of T that can be derived from the Theorems in the former sections.

Theorem 8 *Assume that ρ_0 and ρ_1 are bounded ρ-functions that satisfy R1, with $\rho_1 \leq \rho_0$. We assume that F_0 has a strong unimodal density. Then*

(i) $T_{\mathrm{S}}(F_0) = \alpha_{00}$ is the unique minimizer in (8). If F_0 is symmetric with respect to μ_0 we have $\alpha_{00} = \mu_{0.}$

(ii) $T_{\mathrm{MM}}(F_0) = \alpha_{01}$ is the unique minimizer in (10). If F_0 is symmetric with respect to μ_0 we have $\alpha_{01} = \mu_0$.

(iii) The functional $T = (T_{\mathrm{S}}, T_{\mathrm{MM}}, S)$ is weakly continuous at F_0.

(iv) Assume also that R2 holds and that $a_{00} \neq 0$, $a_{01} \neq 0$, $d_0 \neq 0$. Then, $D_0 = \mathrm{E}_{F_0} \dot{\Psi}(z, T(F_0))$ is invertible, $I_{T_{\mathrm{MM}},F_0}(y)$ is given by (44). The influence function $I_{T_{\mathrm{S}},F_0}(y)$ has a similar expression replacing α_{01}, a_{01} and e_{01} by α_{00}, a_{00} and e_{00} respectively.

(v) Under the same assumptions as in (iv), let $\{F_n\}$ be a sequence of random distributions converging weakly to F_0 and satisfying Condition 2 a.s.. Then T is a.s. weakly differentiable at $\{F_n\}$.

(vi) Assume the same conditions as in (iv) and

$$n^{1/2} I_{T,F_0}(F_n) \to_d H. \tag{45}$$

Then

$$n^{1/2}(T(F_n) - T(F_0)) = n^{1/2} I_{T,F_0}(F_n) + o_p(1), \tag{46}$$

and therefore

$$n^{1/2}(T(F_n) - T(F_0)) \to_d H. \tag{47}$$

(vii) Assume the same conditions as in (iv). Let $\{F_n\}$ be the sequence of empirical distributions corresponding to i.i.d. observations u_i with common distribution F_0. Then (45) holds with $H = N(0, V)$ and V given by (27).

If F_0 is symmetric, the asymptotic variance of T_{MM} given by (40) becomes

$$V = \sigma_0^2 \frac{\mathrm{E}_{F_0} \psi_1 \left(u/\sigma_0\right)^2}{\left(\mathrm{E}_{F_0} \psi_1' \left(u/\sigma_0\right)\right)^2}.$$

7 Proofs

Before proving Theorems 1 and 2 we need some auxiliary results.

Lemma 2 *Consider distributions $\{G_n\}$ and G_0 on $R^p \times R$. Let $\{\xi_n\}$ and $\{\sigma_n\}$ be sequences in $B \times R$ and R_+ respectively, such that $\xi_n \to \xi \in B \times R$ and $\sigma_n \to \sigma > 0$. Assume that $\underline{g}(x,\xi)$ is continuous in ξ. If $G_n \to_w G_0$, then*

$$\lim_{n\to\infty} \mathrm{E}_{G_n} \rho \left(\frac{y - \underline{g}(x,\xi_n)}{\sigma_n} \right) = \mathrm{E}_{G_0} \rho \left(\frac{y - \underline{g}(x,\xi)}{\sigma} \right).$$

Proof. Since $G_n \to_w G_0$ and ρ is continuous and bounded, we have

$$\mathrm{E}_{G_n} \rho \left(\frac{y - \underline{g}(x,\xi)}{\sigma} \right) \to \mathrm{E}_{G_0} \rho \left(\frac{y - \underline{g}(x,\xi)}{\sigma} \right),$$

and therefore it suffices to show that

$$\mathrm{E}_{G_n} \rho \left(\frac{y - \underline{g}(x,\xi_n)}{\sigma_n} \right) - \mathrm{E}_{G_n} \rho \left(\frac{y - \underline{g}(x,\xi)}{\sigma} \right) \to 0.$$

Since $\{G_n\}_{n\geq 1}$ is tight, it suffices to show that if $\mathcal{P}$ is a tight set of distributions of (x,y), then

$$\sup_{F\in\mathcal{P}} \left| \mathrm{E}_F \rho \left(\frac{y - \underline{g}(x,\xi_n)}{\sigma_n} \right) - \mathrm{E}_F \rho \left(\frac{y - \underline{g}(x,\xi)}{\sigma} \right) \right| \to 0.$$

To prove this, put $z = (x,y)$. Then for all $K > 0$

$$\begin{aligned} &\left| \mathrm{E}_F \rho \left(\frac{y - \underline{g}(x,\xi_n)}{\sigma_n} \right) - \mathrm{E}_F \rho \left(\frac{y - \underline{g}(x,\xi)}{\sigma} \right) \right| \qquad (48) \\ &\leq 2\mathrm{E}_F \mathbf{1}_{\{\,||z||>K\}} + \mathrm{E}_F \left| \rho \left(\frac{y - \underline{g}(x,\xi_n)}{\sigma_n} \right) - \rho \left(\frac{y - \underline{g}(x,\xi)}{\sigma} \right) \right| \mathbf{1}_{\{\,||z||\leq K\}}. \end{aligned}$$

If $||z|| \le K$ we have

$$\left| \frac{y - \underline{g}(x,\xi_n)}{\sigma_n} - \frac{y - \underline{g}(x,\xi)}{\sigma} \right|$$
$$\le \frac{1}{\sigma\sigma_n}[|\sigma_n - \sigma||y| + |\sigma_n - \sigma||\underline{g}(x,\xi)| + \sigma|\underline{g}(x,\xi_n) - \underline{g}(x,\xi)|]. \tag{49}$$

Now, given $\varepsilon > 0$, we can find K such that

$$2 \sup_{F \in \mathcal{P}} \mathrm{P}_F\left(||z|| > K\right) \le \varepsilon/2,$$

and α such that

$$|\rho(u) - \rho(v)| \le \varepsilon/2 \text{ if } |u - v| \le \alpha.$$

Then, we can choose n_0 such that the right–hand side of (49) is smaller than α if $n \ge n_0$ and $||z|| \le K$, and so from (48) we obtain for all $n \ge n_0$

$$\left| \mathrm{E}_F \rho\left(\frac{y - \underline{g}(x,\xi_n)}{\sigma_n}\right) - \mathrm{E}_F \rho\left(\frac{y - \underline{g}(x,\xi)}{\sigma}\right) \right| \ \le\ \varepsilon,\ \forall F \in \mathcal{P}.$$

Lemma 3 *Assume that B is closed and let G_0 be any distribution for (x, y) such that (10) has a unique solution $T_{\mathrm{M}}(G_0)$. Let $\{G_n\}$ be a sequence such that $G_n \to_w G_0$ and $\{T_{\mathrm{M}}(G_n)\}$ is bounded. If $\widetilde{S}(G_n) \to \widetilde{S}(G_0) > 0$ then $T_{\mathrm{M}}(G_n) \to T_{\mathrm{M}}(G_0)$.*

Proof Put for brevity

$$\xi_n = T_{\mathrm{M}}(G_n),\ \xi_0 = T_{\mathrm{M}}(G_0),\ \sigma_n = \widetilde{S}(G_n),\ \sigma_0 = \widetilde{S}(G_0). \tag{50}$$

Since $\{\xi_n\}$ remains in a compact set, it suffices to prove that ξ_0 is the only accumulation point of $\{\xi_n\}$. i.e., if a subsequence tends to some $\widehat{\xi}$, then $\widehat{\xi} = \xi_0$. Without loss of generality assume that $\xi_n \to \widehat{\xi}$. The definition of ξ_n implies

$$\mathrm{E}_{G_n} \rho\left(\frac{y - \underline{g}(x,\xi_n)}{\sigma_n}\right) \le \mathrm{E}_{G_n} \rho\left(\frac{y - \underline{g}(x,\xi_0)}{\sigma_n}\right). \tag{51}$$

Using Lemma 2 we get

$$M_{G_0}(\widehat{\xi}) = \mathrm{E}_{G_0} \rho\left(\frac{y - \underline{g}(x,\widehat{\xi})}{\sigma_0}\right) \le \mathrm{E}_{G_0} \rho\left(\frac{y - \underline{g}(x,\xi_0)}{\sigma_0}\right) = M_{G_0}(\xi_0).$$

Since ξ_0 is the only minimizer of M_{G_0}, we conclude that $\widehat{\xi} = \xi_0$.

Lemma 4 *Let $\{\xi_n\}$ and $\{\sigma_n\}$ be sequences in R^{p+1} and R_+, respectively. Assume that when $n \to \infty$, $G_n \to_w G_0$, $||\xi_n|| \to \infty$ and $\{\sigma_n\}$ is bounded. Then*

$$\liminf_{n\to\infty} \mathrm{E}_{G_n} \rho\left(\frac{y - \xi_n'(x', 1)'}{\sigma_n}\right) \ge 1 - c_0, \tag{52}$$

where $c_0 = c\,(G_0)$ is defined in (13).

Proof. Assume without loss of generality that there exist $\gamma \in R^p$ and $\sigma > 0$ such that for some subsequence $\gamma_n = \xi_n/||\xi_n|| \to \gamma$, and $\sigma_n \leq \sigma$. Put $\lambda_n = ||\xi_n||$.

For $\varepsilon > 0$ let d_ε be such that $\rho(u) \geq 1-\varepsilon$ for $|u| \geq d_\varepsilon$. Therefore,

$$\mathrm{E}_{G_n}\rho\left(\frac{y-\xi_n'(x',1)'}{\sigma_n}\right) \geq \mathrm{E}_{G_n}\rho\left(\frac{y-\xi_n'(x',1)'}{\sigma}\right) \geq (1-\varepsilon)\mathrm{P}_{G_n}\left(\frac{|y-\lambda_n\gamma_n'(x',1)'|}{\sigma} \geq d_\varepsilon\right).$$

Then, to prove the Lemma, it suffices to show that

$$\liminf_{n\to\infty} \mathrm{P}_{G_n}\left(\left|\frac{y}{\lambda_n} - \gamma_n'(x',1)'\right| \geq \frac{d_\varepsilon \sigma}{\lambda_n}\right) \geq 1 - c_0.$$

Let $(x_n, y_n) \sim G_n$ and $(x_0, y_0) \sim G_0$. Since $\lambda_n \to \infty$, we have $y_n/\lambda_n \to_p 0$. Then the convergence of γ_n to γ guarantees that

$$\frac{y_n}{\lambda_n} - \gamma_n'(x_n',1)' \to_d \gamma'(x_0',1)'.$$

For any $\alpha > 0$ which is a point of continuity of the distribution of $|\gamma'(x_0,1)|$, $\lambda_n \to \infty$ implies

$$\liminf_{n\to\infty} \mathrm{P}_{G_n}\left(\left|\frac{y}{\lambda_n} - \gamma_n'(x',1)'\right| > \frac{d_\varepsilon \sigma}{\lambda_n}\right) \geq \liminf_{n\to\infty} \mathrm{P}_{G_n}\left(\left|\frac{y}{\lambda_n} - \gamma_n'(x',1)'\right)| > \alpha\right) = \mathrm{P}_{G_0}\left(|\gamma'(x',1)'| > \alpha\right).$$

Letting $\alpha \to 0$ and recalling (13) we get

$$\liminf_{n\to\infty} \mathrm{P}_{G_n}\left(\left|\frac{y}{\lambda_n} - \gamma_n'(x',1)'\right| > \frac{d_\varepsilon \sigma}{\lambda_n}\right) \geq 1 - c_0.$$

The proof of the following Lemma is similar to the one of Lemma 4.

Lemma 5 *Let $\{\xi_n\}$ be a sequence in $B \times R$, with B compact. Assume that when $n \to \infty$, $G_n \to_w G_0$, $||\xi_n|| \to \infty$ and $\{\sigma_n\}$ is bounded. Then*

$$\liminf_{n\to\infty} \mathrm{E}_{G_n}\rho\left(\frac{y - \underline{g}(x,\xi_n)}{\sigma_n}\right) = 1. \tag{53}$$

Finally, the following result we be used.

Lemma 6 *Let $S(G)$ be defined by (9) and suppose that $S(G_0) > 0$. Then, $G_n \to_w G_0$ implies that there exists n_0 such that $S(G_n) > 0$ for $n \geq n_0$.*

Proof: Suppose that the Lemma is not true. Then there exists a subsequence $\{G_{n_k}\}_{k\geq 1}$such that $S(G_{n_k}) = 0$ for all k. This means that giving $\varepsilon > 0$, there exists $(\beta_{n_k}, \alpha_{n_k})$ such that

$$\mathrm{E}_{G_{n_k}}\rho_0\left(\frac{y - g(\mathbf{x}, \beta_{n_k}) - \alpha_{n_k}}{\varepsilon}\right) < \delta \quad \text{for any } s > 0\ .$$

The same arguments that we use to prove Lemma 4 let us show that $\{(\beta_{n_k}, \alpha_{n_k})\}$ is bounded and therefore (passing to a subsequence if necessary) we can assume that $(\beta_{n_k}, \alpha_{n_k}) \to (\tilde{\beta}, \tilde{\alpha})$. Then, from Lemma 2 we get that

$$\mathrm{E}_{G_0}\rho_0\left(\frac{y - g(\mathbf{x}, \tilde{\beta}) - \tilde{\alpha}}{\varepsilon}\right) \leq \delta \quad \text{for any } s > 0.$$

Then, $S(G_0) \leq S^*(G_0, \tilde{\beta}, \tilde{\alpha}) \leq \varepsilon$. Since this holds for any $\varepsilon > 0$, we get that $S(G_0) = 0$. This contradicts the assumption that $S(G_0) > 0$.

7.1 Proof of Theorem 1

Let $G_n \rightarrow_w G_0$. Since $\widetilde{S}$ is weakly continuous at G_0, it follows that $\widetilde{S}(G_n) \rightarrow \widetilde{S}(G_0) > 0$, by hypothesis.

Case (a): We prove first that $\{T_{\mathrm{M}}(G_n)\}$ is bounded. Suppose that it is not true; then without loss of generality we may assume that $||T_{\mathrm{M}}(G_n)|| \rightarrow \infty$. Then Lemma 5 implies

$$1 \ = \ \liminf_{n\rightarrow\infty} M_{G_n}\left(T_{\mathrm{M}}(G_n)\right) \ \leq \ \liminf_{n\rightarrow\infty} M_{G_n}\left(T_{\mathrm{M}}(G_0)\right) = M_{G_0}\left(T_{\mathrm{M}}(G_0)\right),$$

and this implies that $M_{G_0}(\xi) = 1$ for all ξ. This contradicts the assumption that $T_M(G_0)$ is univocally defined. Then, $\{T_{\mathrm{M}}(G_n)\}$ is bounded and from Lemma 3, we get that $T_{\mathrm{M}}(G_n) \rightarrow T_{\mathrm{M}}(G_0)$.

Case (b): Recall the notation in (50). Convergence of $\{\sigma_n\}$ guarantees that it is a bounded sequence. Suppose that $\{\xi_n\}$ is unbounded. Then, passing on to a subsequence if necessary, we may assume that $||\xi_n|| \rightarrow \infty$. In this case by Lemma 4 we have

$$\lim\inf_{n\rightarrow\infty} M_{G_n}(\xi_n) = \lim\inf_{n\rightarrow\infty} \mathrm{E}_{G_n}\rho\left(\frac{y-\xi_n^{'}(x',1)'}{\sigma_n}\right) \geq 1 - c_0. \tag{54}$$

We also have

$$\lim_{n\rightarrow\infty} M_{G_n}(\xi_0) = \lim_{n\rightarrow\infty} \mathrm{E}_{G_n}\rho\left(\frac{y-\xi_0^{'}(x',1)'}{\sigma_n}\right) = M_{G_0}(\xi_0) < 1 - c_0. \tag{55}$$

Inequalities (54) and (55) imply that there exists n_0 such that for $n \geq n_0$

$$M_{G_n}(\xi_n) > M_{G_n}(\xi_0),$$

contradicting the definition of $T_{\mathrm{M}}(G_n)$. Therefore $\{\xi_n\}$ is bounded, and then the weak continuity of T_{M} follows from Lemma 3.

7.2 Proof of Theorem 2

Let $G_n \rightarrow_w G_0$, $\xi_n = T_{\mathrm{S}}(G_n)$, $\xi_0 = T_{\mathrm{S}}(G_0)$, $\sigma_n = S(G_n)$ and $\sigma_0 = S(G_0)$. We prove first that $\{\sigma_n\}$ is bounded. Take any $\sigma_1 > \sigma_0$; then by Lemma 2

$$\mathrm{E}_{G_n}\rho_0\left(\frac{y-\underline{g}(x,\xi_0)}{\sigma_1}\right) \rightarrow \mathrm{E}_{G_0}\rho_0\left(\frac{y-\underline{g}(x,\xi_0)}{\sigma_1}\right) < \delta,$$

and therefore there exists n_0 such that

$$S^*(\xi_0, G_n) < \sigma_1 \text{ for } n \geq n_0, \tag{56}$$

which implies that $S^*(G_n, \xi_0)$ is bounded and therefore $\sigma_n \leq S^*(\xi_0, G_n)$ is also bounded.

On the other hand, by Lemma 6, we get that $\sigma_n > 0$ for n large enough.

We now prove that $\{\xi_n\}$ is bounded. In case (a) if $\{\xi_n\}$ is unbounded, Lemma 5 implies

$$\lim\inf_{n\rightarrow\infty} \mathrm{E}_{G_n}\rho_0\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) \geq 1, \tag{57}$$

and this contradicts the fact that for all n

$$\mathrm{E}_{G_n}\rho_0\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) = \delta < 1.$$

Consider now case (b) and assume that $\{\xi_n\}$ is unbounded . Then, passing on to a subsequence if necessary, we may assume that $||\xi_n|| \to \infty$. Then by Lemma 4

$$\liminf_{n\to\infty} \mathrm{E}_{G_n}\rho_0\left(\frac{y-\xi_n'(x',1)'}{\sigma_n}\right) \geq 1-c_0,$$

and this contradicts the fact that for all n

$$\mathrm{E}_{G_n}\rho_0\left(\frac{y-\xi_n'(x',1)'}{\sigma_n}\right) = \delta < 1-c_0.$$

Then in case (b) $\{\xi_n\}$ is also bounded.

We now show that $\sigma_n \to \sigma_0$. Suppose that this is not true. By passing on to a subsequence if necessary, we may assume that $\sigma_n \to \sigma^* \neq \sigma_0$ and $\xi_n \to \xi^*$ for some ξ^* and σ^*. Since (56) holds for any $\sigma' > \sigma_0$ we have $\sigma^* \leq \sigma_0$ and therefore $\sigma^* < \sigma_0$. Then Lemma 2 implies

$$\delta = \lim_{n\to\infty} \mathrm{E}_{G_n}\rho_0\left(\frac{y-\underline{g}(\xi_n,x)}{\sigma_n}\right) = \mathrm{E}_{G_0}\rho_0\left(\frac{y-\underline{g}(\xi^*,x)}{\sigma^*}\right),$$

and therefore $S\,(G_0) \leq S^*(G_0,\xi^*) = \sigma^* < \sigma_0$. This contradicts the fact that $S(G_0) = \sigma_0$ and shows that S is weakly continuous.

Finally the weak continuity of T_{S} follows from (12) and Theorem 1.

7.3 Proofs of Theorems 3 and 4

The following auxiliary result is due to Ibragimov (1956)

Theorem 9 *If f is a strongly unimodal density and φ is a density such that $\log\varphi$ is concave on its support, the convolution*

$$h(t) = \int_{-\infty}^{\infty} \varphi(u-t)f(u)du \tag{58}$$

is strongly unimodal.

7.3.1 Proof of Theorem 3

(a) Put $k = \int_{-m}^{m} \rho(x)dx$ and $\varphi(u) = (1-\rho(u))/k$, which vanishes for $|u| > m$. Then

$$q(t) = 1 - \mathrm{E}_F(1-\rho(u-t)) = 1 - k\mathrm{E}_F\varphi(u-t) = 1 - kh(t),$$

where $h(t)$ is given by (58). Since by Theorem 9 $h(t)$ is a strongly unimodal density, part (a) of the Theorem follows

(b) It is proved in Lemma 3.1 of Yohai (1985).

7.3.2 Proof of Theorem 4

Without loss of generality we may assume $\sigma = 1$. To prove the Theorem we will show that the unique minimum of $R(\beta,\alpha) = \mathrm{E}_{G_0}\rho(y - g(x,\beta) - \alpha)$ is $\beta = \beta_0, \alpha = t_0$. We will first prove that

$$R(\beta_0, t_0) < R(\beta_0, \alpha) \text{ for } \alpha \neq t_0.$$

This is equivalent to

$$\mathrm{E}_{F_0}\rho(u-t_0) < \mathrm{E}_{F_0}\rho(u-\alpha) \text{ for } \alpha \neq t_0,$$

which follows from Theorem 3.

Consider now (β,α) with $\beta \neq \beta_0$. Let $A = \{x : g(x,\beta_0) = g(x,\beta) + \alpha - t_0\}$ and q as in (15), with F replaced by F_0. Then

$$\begin{aligned} R(\beta,\alpha) &= \mathrm{E}_{G_0}\{\mathrm{E}_{G_0}[\rho(y-g(x,\beta)-\alpha)|x]\} \\ &= \mathrm{E}_{G_0}\{\mathrm{E}_{G_0}[\rho(u+g(x,\beta_0)-g(x,\beta)-\alpha)|x]\}. \end{aligned} \tag{59}$$

Since u and x are independent we get

$$\mathrm{E}[\rho(u+g(x,\beta_0)-g(x,\beta)-\alpha)|x] = q(g(x,\beta)-g(x,\beta_0)+\alpha). \tag{60}$$

Then according to Theorem 3, the left–hand side of (60) is equal to $q(t_0)$ if $x \in A$ and grater than $q(t_0)$ otherwise. The identifiability condition (2) implies that $\mathrm{P}(A^c) > 0$ and from (59) we get that $R(\beta,\alpha) > q(t_0)$. Finally, the Theorem follows from the fact that $R(\beta_0,t_0) = q(t_0)$.

7.4 Proof of Theorems 5 and 6

7.4.1 Proof of Theorem 5

Since

$$\mathrm{E}_{G_n}\Psi(z,T(G_n)) = 0,$$

the Mean Value Theorem together with Condition 2 and the consistency of $T(G_n)$ yield

$$\mathrm{E}_{G_n}\Psi(z,T(G_0)) + D(G_n,\theta_n^*)(T(G_n)-T(G_0)) = 0,$$

where $\theta_n^* \to \theta_0$. Then, (25) implies that $D(G_n,\theta_n^*) \to D_0$ and, since for large n, $D(G_n,\theta_n^*)$ is nonsingular, we may write

$$\begin{aligned} T(G_n) - T(G_0) &= -D(G_n,\theta_n^*)^{-1}\,\mathrm{E}_{G_n}\Psi(z,T(G_0)) \\ &= \mathrm{E}_{G_n} I_{T,G_0}(z) + \left(D_0^{-1} - D(G_n,\theta_n^*)^{-1}\right)\mathrm{E}_{G_n} I_{T,G_0}(z). \end{aligned}$$

Condition 1 implies that the second term of the right–hand side divided by $||\mathrm{E}_{G_n} I_{T,G_0}(z)||$ tends to zero, and this proves the Theorem.

7.4.2 Proof of Theorem 6

Under the assumptions of this Theorem, we can prove that Condition 1 holds a.s. using the same arguments as in Lemma 4.2 of Yohai (1985). The only change is to replace the Law of Large Numbers for i.i.d. random variables by the assumption that $\mathrm{E}_{G_n}d(z) \to \mathrm{E}_{G_0}d(z)$ a.s. for all d such that $\mathrm{E}_{G_0}|d(z)| < \infty$ in the case (a) and for the fact that $\mathrm{E}_{G_n}d(z) \to \mathrm{E}_{G_0}d(z)$ for all function d bounded and continuous in case (b). Then, Theorem 5 implies that T is weakly differentiable at $\{G_n\}$.

7.5 Derivations of influence functions

7.5.1 Derivation of (38)-(39)

Put for brevity

$$t_{\mathrm{MM}} = \frac{y - \underline{g}(x, \xi_{\mathrm{MM}})}{\sigma},\ t_{\mathrm{S}} = \frac{y - \underline{g}(x, \xi_{\mathrm{S}})}{\sigma}.$$

Then

$$\dot{\Psi}(z,\theta) = \begin{bmatrix} \dot{\Psi}_{11}(z,\theta) & 0 & \dot{\Psi}_{13}(z,\theta) \\ 0 & \dot{\Psi}_{22}(z,\theta) & \dot{\Psi}_{23}(z,\theta) \\ \dot{\Psi}_{31}(z,\theta) & 0 & \dot{\Psi}_{33}(z,\theta) \end{bmatrix},$$

where

$$\begin{aligned} \dot{\Psi}_{11}(z,\theta) &= -\frac{1}{\sigma}\psi_0'(t_{\mathrm{S}})\,\underline{\dot{g}}(x,\xi_{\mathrm{S}})\underline{\dot{g}}(x,\xi_{\mathrm{S}})' + \psi_0(t_{\mathrm{S}})\,\underline{\ddot{g}}(x,\xi_{\mathrm{S}}) \\ \dot{\Psi}_{13}(z,\theta) &= -\frac{1}{\sigma}\psi_0'(t_{\mathrm{S}})\,t_{\mathrm{S}}\underline{\dot{g}}(x,\xi_{\mathrm{S}}) \\ \dot{\Psi}_{22}(z,\theta) &= -\frac{1}{\sigma}\psi_1'(t_{\mathrm{MM}})\,\underline{\dot{g}}(x,\xi_{\mathrm{MM}})\underline{\dot{g}}(x,\xi_{\mathrm{MM}})' + \psi_1(t_{\mathrm{MM}})\,\underline{\ddot{g}}(x,\xi_{\mathrm{MM}}) \\ \dot{\Psi}_{23}(z,\theta) &= -\frac{1}{\sigma}\psi_1'(t_{\mathrm{MM}})\,t_{\mathrm{MM}}\underline{\dot{g}}(x,\xi_{\mathrm{MM}}) \\ \dot{\Psi}_{31}(z,\theta) &= -\frac{1}{\sigma}\psi_0(t_{\mathrm{S}})\,\underline{\dot{g}}(x,\xi_{\mathrm{S}}) \\ \dot{\Psi}_{33}(z,\theta) &= -\frac{1}{\sigma}\psi_0(t_{\mathrm{S}})\,t_{\mathrm{S}}. \end{aligned} \tag{61}$$

From (61) it is easy to show that

$$D_0 = \mathrm{E}_{G_0}\dot{\Psi}(z,\theta_0) = -\frac{1}{\sigma_0}\begin{bmatrix} a_{00}C_0 & 0 & e_{00}b_0^* \\ 0 & a_{01}C_0 & e_{01}b_0^* \\ 0 & 0 & d_0 \end{bmatrix}.$$

Therefore $|D_0| = a_{00}a_{01}d_0|C_0|^2$. It follows from (37) $|C_0| \neq 0$ if and on only if $|A_0| \neq 0$, and that

$$C_0^{-1} = \begin{bmatrix} A_0^{-1} & -A_0^{-1}b_0 \\ -(A_0^{-1}b_0)' & 1 + b_0'A_0^{-1}b_0 \end{bmatrix},$$

Direct calculation shows that

$$D_0^{-1} = -\sigma_0\begin{bmatrix} a_{00}^{-1}C_0^{-1} & 0 & -e_{00}a_{00}^{-1}d_0^{-1}C_0^{-1}b_0^* \\ 0 & a_{01}^{-1}C_0^{-1} & -e_{01}a_{01}^{-1}d_0^{-1}C_0^{-1}b_0^* \\ 0 & 0 & d_0^{-1} \end{bmatrix},$$

and the desired results follow from (16).

7.5.2 Derivation of (44)

In this case from (61), it is easy to show that

$$D_0 = -\frac{1}{\sigma_0}\begin{bmatrix} a_{00} & 0 & e_{00} \\ 0 & a_{01} & e_{01} \\ 0 & 0 & d_0 \end{bmatrix},$$

which implies

$$D_0^{-1} = -\sigma_0 \left[\begin{array}{ccc} a_{00}^{-1} & 0 & -e_{00}a_{00}^{-1}d_0^{-1} \\ 0 & a_{01}^{-1} & -e_{01}a_{01}^{-1}d_0^{-1} \\ 0 & 0 & d_0^{-1} \end{array} \right].$$

The rest of the derivation is straightforward.

7.6 Proof of Theorems 7 and 8

7.6.1 Proof of Theorem 7

Part (i) and (ii) follow from Theorem 4 and Remark 2. To prove (iii), we need to check conditions of Theorem 1 and Theorem 2. We start showing that $S(G_0) > 0$. Let

$$h_{\beta,\alpha}(s) = \mathrm{E}\rho_0 \left(\frac{y_i - g(x_i,\beta) - \alpha}{s} \right)$$

Then, we have

$$\lim_{s\to\infty} h_{\beta,\alpha}(s) = \rho_0(0) = 0 \tag{62}$$

and

$$\lim_{s\to 0} h_{\beta,\alpha}(s) = 1 - \mathrm{P}(y_i = g(x_i,\beta) + \alpha). \tag{63}$$

Since u_i has a continuous distribution and is independent of x_i, we also have

$$\mathrm{P}(y_i = g(x_i,\beta) + \alpha) = \mathrm{P}(g(x_i,\beta_0) + u_i = g(x_i,\beta) + \alpha) = \mathrm{E}\left[\mathrm{P}(u_i = g(x_i,\beta) - g(x_i,\beta_0) + \alpha)\right] = 0. \tag{64}$$

Equations (62), (63) and(64) imply that $S^*(G_0,\beta,\alpha) > 0$ for all (β,α), and so $S(G_0) = S^*(G_0,\beta_0,\alpha_{01}) > 0$.

Note that

$$\begin{aligned} M_{G_0}(T_{\mathrm{MM}}(G_0)) &= \mathrm{E}\left(\rho_1\left(\frac{y - T_{\mathrm{MM}}(G_0)}{S(G_0)}\right)\right) \\ &\le \mathrm{E}\left(\rho_1\left(\frac{y - T_{\mathrm{S}}(G_0)}{S(G_0)}\right)\right) \\ &\le \mathrm{E}\left(\rho_0\left(\frac{y - T_{\mathrm{S}}(G_0)}{S(G_0)}\right)\right) \\ &= \delta. \end{aligned}$$

Then $\delta < 1 - C(G_0)$ implies (14) and from Theorem 2 follows that T_{S} and S are weakly continuous. Since S is weekly continuous Theorem 1 implies that T_{MM} is weakly continuous too, and so part (iii) follows.

Part (iv) follows from the formulas obtained in Section 7.5.

(v) follows from part (a) of Theorem (6) while part (vi) follows from Lemma 1. Part (vii) follows from (vi) as was already shown before stating the Theorem.

7.6.2 Proof of Theorem 8

It is completely similar to the proof of Theorem 7. The only differences are that for part (iii) we use that in the case of a location model we have $c\,(G_0) = 0$, and therefore condition (14) reduces to $M_{G_0}(T_{\mathrm{M}}(G_0)) < 1$.

Note that this inequality is implied by the condition that $T_{\mathrm{M}}(G_0)$ is well defined. So, for this case, (14) always holds, and that for part (iv) we use part (b) of Theorem 6 instead of part (a).

Acknowledgements

This research was partially supported by grants X–018 and X–447 from University of Buenos Aires, PID 5505 and PIP 0592 from CONICET and PICTs 21407, 00883 and 00899 from ANPCyT.

References

References

[1] Billingsley, P. (1999). *Convergence of Probability Measures (2nd. ed.). Wiley, New York.*

[2] Boos, D.D. and Serfling, R.J. (1980). A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates. *The Annals of Statistics*, **3**, 616-624.

[3] Cízek, P. (2006). Least trimmed squares in nonlinear regression under dependence. Journal of Statistical Planning and Inference, 136, 3967-3988.

[4] Clarke, B.R..(1983) Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *The Annals of Statistics*, **4**, 1196-1205.

[5] Clarke, B.R. (2000). A remark on robustness and weak continuity of M-estimators. *Journal of the Australian Mathematical Society*, **68**, 411-418.

[6] Fasano, M.V. (2009). Robust estimation in nonlinear regression. Ph. D. Thesis, University of La Plata. Available at http://www.mate.unlp.edu.ar/tesis/tesis_fasano_v.pdf

[7] Fernholz, L.T. (1983). Von Mises Calculus for Statistical Functionals. *Lecture Notes in Statistics* **19**, Springer-Verlag, New York

[8] Fraiman, R. (1983). General M-estimators and applications to bounded influence estimation for non-linear regression, *Communications In Statistics, Theory and Methods*, **12,** 2617-2631.

[9] Hampel, F.R. (1971). A general definition of qualitative robustness, *The Annals of Mathematical Statistics,* **42**, 1887–1896.

[10] Hampel, F.R. (1974).The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, 69, 383–393.

[11] Ibragimov, I. A. (1956). On the Composition of Unimodal Distributions. *Theory of Probability and its Applications,* **1**, 255-260.

[12] Liese, F. and Vajda, I. (2003). A general asymptotic theory of M-estimators I. Mathematical Methods of Statistics, 12, 454-477.

[13] Liese, F. and Vajda, I. (2004). A general asymptotic theory of M-estimators II Mathematical Methods of Statistics, 13, 82-95.

[14] Maronna, R.A., Martin, R. D. and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods.* Springer, London.

[15] Mizera, I. (1993), On Consistent M-Estimators: Tuning Constants, Unimodality and Breakdown. *Kybernetika*, **30**, 289–300.

[16] Papantoni-Kazakos, P. and Grey, R. (1979). Robustness of estimators of stationary observations. *The Annals of Probability*, **6,** 989-1002.

[17] Richardson, G.D: and Bhattacharyya, B.B. (1986). Consistent estimators in nonlinear regression for a noncompact parameter space. The Annals of Statistics, 14, 1591-1596.

[18] Rousseeuw, P.J and Yohai, V.J. (1984), Robust regression by means of S-estimators, *Robust and Nonlinear Time Series*, J. Franke,W. Härdle and R.D. Martin (eds.), Lectures Notes in Statistics 26, 256–272, New York: Springer.

[19] Stromberg, A.J. and Ruppert, D. (1992), "Breakdown in Nonlinear Regression," *Journal of the American Statistical Association*, **87**, 991-997.

[20] Sakata, S. and White, H. (2001). S-estimation of nonlinear regression models with dependent and heterogeneous observations. Journal of Econometrics, 103, 5-72.

[21] Stromberg, A. J. (1995). Consistency of the least median of squares estimator in nonlinear regression Communications in Statistics: Theory and Methods, 24, 1971-1984.

[22] Vainer, B. P. and Kukush, A. G. (1998). The consistency of M-estimators constructed from a concave weight function. Theory of Probability and Mathematical Statistics, 57, 11-18.

[23] Yohai, V.J. (1985).High Breakdown Point and High Efficiency Robust Estimates for Regression. Technical report No. 66. Department of Statistics, University of Washington. Available at http://www.stat.washington.edu/research/reports/1985/tr066.pdf.

[24] Yohai, V.J. (1987), High breakdown–point and high efficiency estimates for regression, *The Annals of Statistics* **15**, 642–65.